WHAT IS CLAIMED IS:

1. In a computer system having a large disk resident data set, a method of analyzing the disk resident data set using a patient rule induction method (PRIM), comprising steps of:

(a) receiving a meta parameter and a relational data table comprised of continuous attributes, discrete attributes and a cost attribute, wherein the cost attribute represents cost output values based on continuous attributes values and discrete attribute values as inputs;

(b) defining a hyper-rectangle enclosing a multi-dimensional space defined by the continuous attribute values and the discrete attribute values, wherein the continuous attribute values and the discrete attribute values are represented as points within the multi-dimensional space;

(c) removing a plurality of points along edges of the hyper-rectangle based on an average of the cost output value from the plurality of points until a count of the points enclosed within the hyper-rectangle equals the meta parameter; and

(d) adding removed discrete attribute value points and the continuous attribute value points along the edges of the hyper-rectangle until a sum of the cost output values over the multi-dimensional space enclosed by the hyper-rectangle changes.

2. The method of claim 1, wherein the continuous attributes are A_c continuous attributes, the discrete attributes are A_d discrete attributes and the meta parameter is β_0, step (b) further including steps of:

(b)(i) separating the data into $A_c + 2$ lists, such that a list is generated for each continuous attribute to form A_c continuous attribute lists containing the continuous attribute values, a discrete attribute list containing the A_d discrete attributes and the discrete attribute values and a cost attribute list containing the cost output values;

(b)(ii) adding a label to each of the A_c continuous attribute lists, the discrete attribute list and the cost attribute list, such that the label is an index of a tuple to which the respective attribute value belongs within the relational data table;

(b)(iii) sorting the A_c continuous attribute lists based on a continuous attribute value in each row of the A_c continuous attribute lists; and

(b)(iv) adding a label to the cost list, such that the label is a cost flag that indicates whether the tuple containing the cost output value is enclosed within the hyper-rectangle, such that the cost flag is initially set to one.

3. The method of claim 2, wherein, step (c) further including steps of:

(c)(i) determining the discrete attribute value enclosed within the plurality of points along an edge of the hyper-rectangle with a lowest average cost output value;

(c)(ii) determining the continuous attribute value enclosed within the plurality of points along an edge of the hyper-rectangle with a lowest average cost output value;

(c)(iii) comparing the lowest average cost output value determined in step (c)(i) with the lowest average cost output value determined in step (c)(ii) to determine an attribute with the lowest average cost output value; and

(c)(iv) removing all continuous attribute value points and all discrete attribute value points of the tuples containing the attribute determined in step (c)(iii) from the hyper-rectangle; and

(c)(v) repeating steps (c)(i) to (c)(iv) until the count of the points within the hyper-rectangle equals β_0.

4. The method of claim 3, wherein, step (c)(i) further including steps of:

1) generating A_d discrete histograms, one for each discrete attribute containing the discrete attribute value and an average of the cost output value for each tuple containing the discrete attribute value; and

2) comparing the average cost output value for each discrete attribute value to determine the discrete attribute value with the lowest average cost output value.

5. The method of claim 4, wherein step 1) further includes a step of:

assigning a code to the discrete attribute list, wherein the discrete attribute list is sorted based on the assigned code in order to optimize step (c), such that the discrete attribute values are ground together according to their discrete attribute.

6. The method of claim 3, wherein, each continuous attribute list is sorted in increasing order with a start pointer to a first row in each continuous attribute list, and a second continuous attribute list is sorted in decreasing order with an end pointer to a first row in each second continuous attribute list, step (c)(ii) further including steps of:

1) marking a start cutoff value in each of the A_c continuous attribute lists based on a count of the tuples containing the discrete attribute value determined in step (c)(i) and enclosed within the hyper-rectangle;

2) marking an end cutoff value in each of the second continuous attribute lists based on the count of the tuples containing the discrete attribute value determined in step (c)(i) and enclosed within the hyper-rectangle;

3) determining a start cost average value for each continuous output value between the start pointer and the cutoff value for each of the A_c cost attribute lists;

4) determining an end cost average value for each continuous output value between the cutoff value and the end pointer for each of the second cost attribute lists;

5) generating A_c continuous histograms, each containing the continuous attribute and the average cost output value, wherein the average cost output value is a lesser of the start cost average value and the end cost average value; and

6) comparing the average cost output value for each continuous histogram to determine the continuous attribute value with the lowest average cost output value.

7. The method of claim 6, wherein the discrete attribute list is sorted based on an assigned code thereby grouping the discrete attribute values according to their discrete attribute, step (c)(iv) further includes steps of:

when the attribute determined in step (c)(iii) is a continuous attribute,

1) when the start cost average value is less than the end cost average value, setting the cost flag equal to zero for each continuous attribute value between the start pointer and the start cutoff value using the index of the continuous attribute value to reference the cost flag,

2) setting the start pointer equal to the start cutoff value;

3) when the end cost average value is less than the start cost average value, setting the cost flag equal to zero for each continuous attribute value between the end cutoff

value and the end pointer the using the index of the continuous attribute value to reference the cost flag,

4) setting the end pointer equal to the end cutoff value;

when the attribute determined in step (c)(iii) is a discrete attribute; and

5) setting the cost flag equal to zero for each discrete attribute value equal to the attribute determined in step (c)(iii) using the index of the discrete attribute value to reference the cost flag.

8. The method of claim 2, wherein a total cost output is a sum of the cost output values over the multidimensional space enclosed by the hyper-rectangle following step (c), and the cost attribute list includes a cost counter, step (d) further including steps of:

(d)(i) for each tuple with the cost flag set to zero, incrementing the cost counter for each point belonging to the tuple that is not enclosed with the hyper-rectangle;

(d)(ii) determining the discrete attribute value outside of the points enclosed by the hyper-rectangle with a highest average cost output value;

(d)(iii) determining the continuous attribute value outside of the points enclosed by the hyper-rectangle with a highest average cost output value;

(d)(iv) comparing the highest average cost output value determined in step (d)(ii) with the highest average cost output value determined in step (d)(iii) to determine an attribute with the highest average cost output value;

(d)(v) decrementing the cost counter for all continuous attribute value points and all discrete attribute value points belonging to the tuples containing the attribute determined in step (d)(iv), such that attributes with the cost counter equal to zero are enclosed within the hyper-rectangle; and

(d)(vi) repeating steps (d)(i) to (d)(v) until a sum of the cost output value over the plurality of points enclosed by the hyper-rectangle is less than the total cost output.

9. The method of claim 8, wherein, step (d)(ii) further including steps of:

1) generating A_d discrete histograms, one for each discrete attribute containing the discrete attribute value and an average of the cost output value for each tuple containing the discrete attribute value; and

2) comparing the average cost output value for each discrete attribute value to determine the discrete attribute value with the highest average cost output value.

10. The method of claim 8, wherein the A_c continuous attribute lists are sorted in increasing order and each contain a start pointer and an end pointer, such that the continuous attribute values there between are enclosed within the hyper-rectangle, step (d)(iii) further including steps of:

1) sorting the continuous attribute values in the A_c continuous attributes lists between a first row in the A_c continuous attributes lists and the start pointer in decreasing order;

2) marking a start cutoff value in each of the A_c continuous attribute lists based on a count of the tuples containing the discrete attribute value determined in step (d)(ii) and enclosed within the hyper-rectangle;

3) marking an end cutoff value in each of the A_c continuous attribute lists based on the count of the tuples containing the discrete attribute value determined in step (d)(ii) and enclosed within the hyper-rectangle;

4) determining a start cost average value for each continuous output value between the start pointer and the cutoff value for each of the A_c continuous attribute lists;

5) determining an end cost average value for each continuous output value between the end pointer and the cutoff value for each of the A_c continuous attribute lists;

6) generating A_c continuous histograms, each containing the continuous attribute and the average cost output value, wherein the average cost output value is a greater of the start cost average value and the end cost average value; and

7) comparing the average cost output value for each continuous histogram to determine the continuous attribute value with the highest average cost output value.

11. The method of claim 10, wherein the discrete attribute list is sorted based on an assigned code thereby grouping the discrete attribute values according to their discrete attribute, step (d)(v) further includes steps of:

when the attribute determined in step (d)(iv) is a continuous attribute,

1) when the start cost average value is less than the end cost average value, decrementing the cost counter for each continuous attribute value between the start cutoff value and the start pointer using the index of the continuous attribute value to reference the cost counter;

2) setting the start pointer equal to the start cutoff value;

when the attribute determined in step (d)(iv) is a discrete attribute; and

3) decrementing the cost counter for each discrete attribute value equal to the attribute determined in step (d)(iv) using the index of the discrete attribute value to reference the cost counter.

12. In a parallel architecture computer system having a large disk resident data set, a method of analyzing the disk resident data set in parallel using a patient rule induction method (PRIM), comprising steps of:

(a) receiving a relational table of data comprised of A_c continuous attributes, A_d discrete attributes, a meta parameter β_0 and a cost attribute, wherein the cost attribute represents cost output values based on continuous attributes values and discrete attribute values as inputs;

(b) defining a hyper-rectangle enclosing a multi-dimensional space defined by the continuous attribute values and the discrete attribute values, wherein the continuous attribute values and the discrete attribute values are represented as points within multi-dimensional space;

(c) separating the data into $A_c + 2$ lists, such that a list is generated for each continuous attribute to form A_c continuous attribute lists containing the continuous attribute values, a discrete attribute list containing the A_d discrete attributes and the discrete attribute values and a cost attribute list containing the cost output values;

(d) sorting the A_c continuous attribute lists in parallel among a plurality of processors based on a continuous attribute value in each row of the A_c continuous attribute lists;

(e) striping the A_c continuous attribute lists and the discrete attribute lists across the plurality of processor, wherein each processor contains a copy of the cost attribute list;

(f) removing a plurality of points along edges of the hyper-rectangle using reduction and a one to all broadcast based on an average of the cost output value from the plurality of

points until a count of the points enclosed within the hyper-rectangle equals the meta parameter; and

(g) adding removed discrete attribute value points and the continuous attribute value points along the edges of the hyper-rectangle using reduction and a one to all broadcast until a sum of the cost output values over the multi-dimensional space enclosed by the hyper-rectangle changes.

13. The method of claim 12, wherein step (c) further including steps of:

(c)(i) adding a label to each of the A_c continuous attribute lists, the discrete attribute list and the cost attribute list, such that the label is an index of a tuple to which the respective attribute belongs within the relational data table;

(c)(ii) adding a label to the cost list, such that the label is a cost flag that indicates whether the tuple containing the cost output value is enclosed within the hyper-rectangle, such that the cost flag is initially set to one.

14. The method of claim 13, wherein, step (f) further including steps of:

(f)(i) determining the discrete attribute value enclosed within the plurality of points along an edge of the hyper-rectangle with a lowest average cost output value using reduction;

(f)(ii) determining the continuous attribute value enclosed within the plurality of points along an edge of the hyper-rectangle with a lowest average cost output value using reduction;

(f)(iii) comparing the lowest average cost output value determined in step (f)(i) with the lowest average cost output value determined in step (f)(ii) to determine an attribute with the lowest average cost output value;

(f)(iv) removing all continuous attribute value points and all discrete attribute value points of the tuples containing the attribute determined in step (c)(iii) from the hyper-rectangle by setting the cost flag to zero using the index of the attribute to reference the cost flag in the cost list contained in each of the plurality of processors using the one to all broadcast; and

(f)(v) repeating steps (f)(i) to (f)(iv) until the count of the points within the hyper-rectangle equals β_0.

15. The method of claim 14, wherein a total cost output is a sum of the cost output values over the multidimensional space enclosed by the hyper-rectangle following step (f), and the cost attribute list includes a cost counter, step (g) further including steps of:

(g)(i) for each tuple with the cost flag set to zero, incrementing the cost counter for each point belonging to the tuple that is not enclosed with the hyper-rectangle among each of the plurality of processors;

(g)(ii) determining the discrete attribute value outside of the points enclosed by the hyper-rectangle with a highest average cost output value using reduction;

(g)(iii) determining the continuous attribute value outside of the points enclosed by the hyper-rectangle with a highest average cost output value using reduction;

(g)(iv) comparing the highest average cost output value determined in step (g)(ii) with the highest average cost output value determined in step (g)(iii) and determine which attribute with the highest average cost output value;

(g)(v) decrementing the cost counter for all continuous attribute value points and all discrete attribute value points belonging to the tuples containing the attribute determined in step (g)(iv) using the index of the attribute to reference the cost flag in the cost list contained in each of the plurality of processors using the one to all broadcast, such that attributes with the cost counter equal to zero are enclosed within the hyper-rectangle; and

(g)(vi) repeating steps (g)(i) to (g)(v) until a sum of the cost output value over the plurality of points enclosed by the hyper-rectangle is less than the total cost output.

16. In a symmetric multi-processor architecture computer system having a large disk resident data set, a method of analyzing the disk resident data set in parallel using a patient rule induction method (PRIM), comprising steps of:

(a) receiving a relational table of data comprised of A_c continuous attributes, A_d discrete attributes, a meta parameter β_0 and a cost attribute, wherein the cost attribute represents cost output values based on continuous attributes values and discrete attribute values as inputs;

(b) defining a hyper-rectangle enclosing a multi-dimensional space defined by the continuous attribute values and the discrete attribute values, wherein the continuous attribute values and the discrete attribute values are represented as points within multi-dimensional

space;

(c) separating the data into $A_c + 2$ lists, such that a list is generated for each continuous attribute to form A_c continuous attribute lists containing the continuous attribute values, a discrete attribute list containing the A_d discrete attributes and the discrete attribute values and a cost attribute list containing the cost output values;

(d) sorting the A_c continuous attribute lists in parallel based on a continuous attribute value in each row of the A_c continuous attribute lists;

(e) striping the A_c continuous attribute lists and the discrete attribute lists over a plurality of portions of a shared disk, wherein each portion of the shared disk contains a copy of the cost attribute list;

(f) removing a plurality of points along edges of the hyper-rectangle using reduction and a one to all broadcast based on an average of the cost output value from the plurality of points until a count of the points enclosed within the hyper-rectangle equals the meta parameter; and

(g) adding removed discrete attribute value points and the continuous attribute value points along the edges of the hyper-rectangle using reduction and a one to all broadcast until a sum of the cost output values over the multi-dimensional space enclosed by the hyper-rectangle changes.

17. The method of claim 16, wherein, step (c) further including steps of:

(c)(i) adding a label to each of the A_c continuous attribute lists, the discrete attribute list and the cost attribute list, such that the label is an index of a tuple to which the respective attribute belongs within the relational data table;

(c)(ii) adding a label to the cost list, such that the label is a cost flag that indicates whether the tuple containing the cost output value is enclosed within the hyper-rectangle, such that the cost flag is initially set to one.

18. The method of claim 17, wherein, step (f) further including steps of:

(f)(i) determining the discrete attribute value enclosed within the plurality of points along an edge of the hyper-rectangle with a lowest average cost output value using reduction;

(f)(ii) determining the continuous attribute value enclosed within the plurality of points along an edge of the hyper-rectangle with a lowest average cost output value using reduction;

(f)(iii) comparing the lowest average cost output value determined in step (f)(i) with the lowest average cost output value determined in step (f)(ii) to determine an attribute with the lowest average cost output value;

(f)(iv) removing all continuous attribute value points and all discrete attribute value points of the tuples containing the attribute determined in step (f)(iii) from the hyper-rectangle by setting the cost flag to zero using the index of the attribute to reference the cost flag in the cost list contained in each of the plurality of processors using the one to all broadcast; and

(f)(v) repeating steps (f)(i) to (f)(iv) until the count of the points within the hyper-rectangle equals β_0.

19. The method of claim 18, wherein a total cost output is a sum of the cost output values over the multidimensional space enclosed by the hyper-rectangle following step (f), and the cost attribute list includes a cost counter, step (g) further including steps of:

(g)(i) for each tuple with the cost flag set to zero, incrementing the cost counter for each point belonging to the tuple that is not enclosed with the hyper-rectangle among each of the plurality of portions of the shared disk;

(g)(ii) determining the discrete attribute value outside of the points enclosed by the hyper-rectangle with a highest average cost output value using reduction;

(g)(iii) determining the continuous attribute value outside of the points enclosed by the hyper-rectangle with a highest average cost output value using reduction;

(g)(iv) comparing the highest average cost output value determined in step (g)(ii) with the highest average cost output value determined in step (g)(iii) and determine which attribute with the highest average cost output value;

(g)(v) decrementing the cost counter for all continuous attribute value points and all discrete attribute value points belonging to the tuples containing the attribute determined in step (g)(iv) using the index of the attribute to reference the cost flag in the cost list contained in each of the plurality of processors using the one to all broadcast, such that attributes with

the cost counter equal to zero are enclosed within the hyper-rectangle; and

(g)(vi) repeating steps (g)(i) to (g)(v) until a sum of the cost output value over the plurality of points enclosed by the hyper-rectangle is less than the total cost output.